

The Stability of Low-Rank Matrix Reconstruction: a Constrained Singular Value View

Gongguo Tang and Arye Nehorai

Department of Electrical and Systems Engineering

Washington University in St. Louis

St. Louis, MO 63130-1127

Email: nehorai@ese.wustl.edu

Abstract

The stability of low-rank matrix reconstruction is investigated in this paper. The ℓ_* -constrained minimal singular value (ℓ_* -CMSV) of the measurement operator is shown to determine the recovery performance of nuclear norm minimization based algorithms. Compared with the stability results using the matrix restricted isometry constant, the performance bounds established using ℓ_* -CMSV are more concise and tight, and their derivations are less complex. The computationally amenable ℓ_* -CMSV and its associated error bounds also have more transparent relationships with the Signal-to-Noise Ratio. Several random measurement ensembles are shown to have ℓ_* -CMSVs bounded away from zero with high probability, as long as the number of measurements is relatively large.

Index Terms

ℓ_* -constrained minimal singular value, matrix Basis Pursuit, matrix Dantzig selector, matrix LASSO estimator, restricted isometry property

This work was supported by the Department of Defense under the Air Force Office of Scientific Research MURI Grant FA9550-05-1-0443, and ONR Grant N000140810849.

I. INTRODUCTION

The last decade witnessed the burgeoning of exploiting low dimensional structures in signal processing, most notably the sparseness for vectors [1], [2], low-rankness for matrices [3]–[5], and low-dimensional manifold structure for general non-linear data set [6], [7]. This paper focuses on the stability problem of low-rank matrix reconstruction. Suppose $X \in \mathbb{R}^{n_1 \times n_2}$ is a matrix of rank $r \ll \min\{n_1, n_2\}$, the low-rank matrix reconstruction problem aims to recover matrix X from a set of linear measurements \mathbf{y} corrupted by noise \mathbf{w} :

$$\mathbf{y} = \mathcal{A}(X) + \mathbf{w}, \quad (1)$$

where $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^m$ is a linear measurement operator. Since the matrix X lies in a low-dimensional sub-manifold of $\mathbb{R}^{n_1 \times n_2}$, we expect $m \ll n_1 n_2$ measurements would suffice to reconstruct X from \mathbf{y} by exploiting the signal structure. Application areas of model (1) include factor analysis, linear system realization [8], [9], matrix completion [10], [11], quantum state tomography [12], face recognition [13], [14], Euclidean embedding [15], to name a few (See [3]–[5] for discussions and references therein).

Several considerations motivate the study of the stability of low-rank matrix reconstruction. First, in practical problems the linear measurement operator \mathcal{A} is usually used repeatedly to collect measurement vectors \mathbf{y} for different matrices X . Therefore, before taking the measurements, it is desirable to know the goodness of the measurement operator \mathcal{A} as far as reconstructing X is concerned. Second, a stability analysis would offer means to quantify the confidence on the reconstructed matrix X , especially when there is no other ways to justify the correctness of the reconstructed signal. In addition, as in the case of sparse signal reconstruction [16], in certain applications we have the freedom to design the measurement operator \mathcal{A} by selecting the best one from a collection of operators, which requires a precise quantification of the goodness of any given operator. All these considerations suggest that the stability measure should be *computable*, an aspect usually overlooked in literature.

This work is in parallel with our previous work on the stability study of sparse signal reconstruction. In [16], we demonstrated that the ℓ_1 -constrained minimal singular value (ℓ_1 -CMSV) of a measurement matrix quantifies the stability of sparse signal reconstruction. Several important random measurement ensembles are shown to have ℓ_1 -CMSVs bounded away from

zero with reasonable number of measurements. More importantly, we designed several algorithms to compute the ℓ_1 -CMSV of any given measurement matrix.

In the current work, we define the ℓ_* -constrained minimal singular value (ℓ_* -CMSV) of a linear operator to measure the stability of low-rank matrix reconstruction. A large class of random linear operators are also shown to have ℓ_* -CMSVs bounded away from zero. The analysis for random linear operators acting on matrix space is more challenging. We need to employ advanced tools from geometrical functional analysis and empirical processes. The computational aspect of ℓ_* -CMSV is left to future work.

Several works in literature also address the problem of low-rank matrix reconstruction. Recht *et.al.* study the recovery of X in model (1) in the noiseless setting [3]. The matrix restricted isometry property (mRIP) is shown to guarantee exact recovery of X subject to the measurement constraint $\mathcal{A}(X) = \mathbf{y}$. Candès *et.al.* consider the noisy problem and analyze the reconstruction performance of several convex relaxation algorithms [5]. The techniques used in this paper for deriving the error bounds in terms of ℓ_* -CMSV draw ideas from [5]. Our bounds are more concise and are expected to be tighter. In both works [3] and [5], several important random measurement ensembles are shown to have the matrix restricted isometry constant (mRIC) bounded away from zero for reasonably large m . Our procedures for establishing the parallel results for the ℓ_* -CMSV are significantly different from those in [3] and [5]. By analogy to the ℓ_1 -CMSV, we expect that the ℓ_* -CMSV is computationally more amenable than the mRIC [16].

The ℓ_* -CMSV has several advantages over the mRIC in stability analysis of low-rank matrix analysis. First, the error bounds involving ℓ_* -CMSV have more transparent relationships with the Sign-to-Noise-Ratio. For example, consider the matrix Basis Pursuit algorithm, if we multiply the measurement operator \mathcal{A} by a positive constant, the ℓ_* -CMSV will scale by the same constant and the error bound for the matrix Basis Pursuit will scale inverse proportionally, while the mRIC and associated error bounds have more complex scaling properties. Second, the ℓ_* -CMSV are computationally more amenable than the mRIC. The discrete nature of mRIC makes its computation algorithm design very difficult. In contrast, many tools at our disposal can deal with the continuous formulation of ℓ_* -CMSV, for example, the Lagrange multiplier or the Karush-Kuhn-Tucker condition [17]. In particular, many algorithms for computing the smallest matrix singular values such as the conjugate gradient method for Rayleigh quotient minimization [18], [19] might be adapted to the computation of ℓ_* -CMSV by incorporation one additional constraint.

We will investigate these possibilities in followup works. In addition, the derivation of the ℓ_* -CMSV bounds is less complicated and the resulting bounds have more concise forms. These bounds are also expected to be tighter than those in terms of mRIC.

The paper is organized as follows. Section II previews our main results. Section III introduces notations, the measurement models, three convex relaxation based recovery algorithms, and the definition and properties of mRIC. Section IV is devoted to deriving error bounds in terms of the ℓ_* -CMSV for three convex relaxation algorithms. In Section V, we analyze the ℓ_* -CMSV for isotropic and subgaussian measurement operators. The paper is concluded in Section VI.

II. OVERVIEW OF THE MAIN RESULTS

In this section, we preview the main results of this paper. All proofs are postponed to latter sections. Throughout the paper, we will assume $n_1 \leq n_2$ such that $\min\{n_1, n_2\} = n_1$ and $\max\{n_1, n_2\} = n_2$.

A. ℓ_* -Constrained Singular Values

We first introduce a quantity that continuously extends the concept of rank for a given matrix X . It is also an extension of the ℓ_1 -sparsity level from vectors to matrices [16].

Definition 1 *The ℓ_* -rank of a non-zero matrix $X \in \mathbb{R}^{n_1 \times n_2}$ is defined as*

$$\tau(X) = \frac{\|X\|_*^2}{\|X\|_F^2} = \frac{\|\sigma(X)\|_1^2}{\|\sigma(X)\|_2^2}, \quad (2)$$

where $\|\cdot\|_*$ is the nuclear norm of a matrix and $\|\cdot\|_F$ the Frobenius norm. We use $\sigma(X) \in \mathbb{R}^{n_1}$ to denote the vector of singular values of X in decreasing order.

The scaling invariant $\tau(X)$ is indeed a measure of rank. To see this, suppose $\text{rank}(X) = r$; then Cauchy-Schwarz inequality implies that

$$\tau(X) \leq r, \quad (3)$$

and we have equality if and only if all non-zero singular values of X are equal. Therefore, the more non-zero singular values X has and the more evenly the magnitudes of these non-zero singular values are distributed, the larger $\tau(X)$. In particular, if X is of rank 1, then $\tau(X) = 1$; if X is of full rank n_1 with all singular values having the same magnitudes, then $\tau(X) = n_1$.

However, if X has n_1 non-zero singular values but their magnitudes are spread in a wide range, then its ℓ_* -rank might be very small.

Definition 2 For any $\tau \in [1, n_1]$ and any linear operator $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \mapsto \mathbb{R}^m$, define the ℓ_* -constrained minimal singular value (abbreviated as ℓ_* -CMSV) and the ℓ_* -constrained maximal singular value of \mathcal{A} by

$$\rho_\tau^{\min}(\mathcal{A}) \stackrel{\text{def}}{=} \inf_{X \neq 0, \tau(X) \leq \tau} \frac{\|\mathcal{A}(X)\|_2}{\|X\|_F}, \quad \text{and} \quad (4)$$

$$\rho_\tau^{\max}(\mathcal{A}) \stackrel{\text{def}}{=} \sup_{X \neq 0, \tau(X) \leq \tau} \frac{\|\mathcal{A}(X)\|_2}{\|X\|_F}, \quad (5)$$

respectively. Because we mainly use ρ_τ^{\min} in this paper, for notational simplicity, we sometimes use ρ_τ to denote ρ_τ^{\min} when it causes no confusion.

For an operator \mathcal{A} , a non-zero $\rho_\tau(\mathcal{A}) = \rho_\tau^{\min}(\mathcal{A})$ roughly means that \mathcal{A} is invertible when restricted onto the set $\{X \in \mathbb{R}^{n_1 \times n_2} : \tau(X) \leq \tau\}$, or equivalently, the intersection of the null space of \mathcal{A} and $\{X \in \mathbb{R}^{n_1 \times n_2} : \tau(X) \leq \tau\}$ contains only the null vector of $\mathbb{R}^{n_1 \times n_2}$. The value of $\rho_\tau(\mathcal{A})$ measures the invertibility of \mathcal{A} restricted onto $\{\tau(X) \leq \tau\}$. As we will see in Section IV, the error matrices for convex relaxation algorithms have small ℓ_* -ranks. Therefore, the error matrix is distinguishable from the zero matrix given the image of the error matrix under \mathcal{A} . Put it another way, given noise corrupted $\mathcal{A}(X)$, a signal matrix X is distinguishable from $X + H$, as long as the noise works in a way such that the error matrix H has a small ℓ_* -rank. This explains roughly why $\rho_\tau(\mathcal{A})$ determines the performance of convex relaxation algorithms.

We begin to define a class of important random operator ensembles: the isotropic and subgaussian ensemble, after introducing some prerequisite concepts. For a scalar random variable x , the Orlicz ψ_2 norm is defined as

$$\|x\|_{\psi_2} = \inf \left\{ t > 0 : \mathbb{E} \exp \left(\frac{|x|^2}{t^2} \right) \leq 2 \right\}. \quad (6)$$

Markov's inequality immediately gives that x with finite $\|x\|_{\psi_2}$ has subgaussian tail:

$$\mathbb{P}(|x| \geq t) \leq 2 \exp(-ct^2/\|x\|_{\psi_2}). \quad (7)$$

The converse is also true, i.e., if x has subgaussian tail $\exp(-t^2/K^2)$, then $\|x\|_{\psi_2} \leq cK$.

Definition 3 A random vector $\mathbf{x} \in \mathbb{R}^n$ is called isotropic and subgaussian with constant L if $\mathbb{E} |\langle \mathbf{x}, \mathbf{u} \rangle|^2 = \|\mathbf{u}\|_2^2$ and $\|\langle \mathbf{x}, \mathbf{u} \rangle\|_{\psi_2} \leq L \|\mathbf{u}\|_2$ hold for any $\mathbf{u} \in \mathbb{R}^n$.

A random vector \mathbf{x} with independent subgaussian entries x_1, \dots, x_n is a subgaussian vector because [20]

$$\begin{aligned} \|\langle \mathbf{x}, \mathbf{u} \rangle\|_{\psi_2} &\leq c \left(\sum_{i=1}^n u_i^2 \|x_i\|_{\psi_2}^2 \right)^{1/2} \\ &\leq c \|\mathbf{u}\|_2 \max_{1 \leq i \leq n} \|x_i\|_{\psi_2}. \end{aligned} \quad (8)$$

Clearly, if in addition $\{x_i\}_{i \leq n}$ are centered and has unit variance, then \mathbf{x} is also isotropic. In particular, the standard Gaussian vector on \mathbb{R}^n and the sign vector with *i.i.d.* $1/2$ Bernoulli entries are isotropic and subgaussian. Isotropic and subgaussian random vectors also include the vectors with the normalized volume measure on various convex symmetric bodies, for example, the unit balls of ℓ_p^n for $2 \leq p \leq \infty$ [21].

Clearly, any linear operator $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^m$ can be represented by a collection of matrices $\mathcal{A} = \{A_1, \dots, A_m\}$ (Refer to Section III-A for more details). Based on this representation of \mathcal{A} , we have the following definition of isotropic and subgaussian operators:

Definition 4 Suppose $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^m$ is a linear operator with corresponding matrix representation \mathcal{A} . We say \mathcal{A} is from the isotropic and subgaussian ensemble if for each $A_i \in \mathcal{A}$ $\text{vec}(A_i)$ is an independent isotropic and subgaussian vector with constant L , and L is a numerical constant independent of n_1, n_2 .

Isotropic and subgaussian operators include operators with *i.i.d.* centered subgaussian entries of unit variance (Gaussian and Bernoulli entries in particular) as well as operators whose matrices A_i ($\text{vec}(A_i)$, more precisely) are independent copies of random vectors distributed according to the normalized volume measure of unit balls of $\ell_p^{n_1 n_2}$ for $2 \leq p \leq \infty$.

For any isotropic and subgaussian operator \mathcal{A} the typical values of $\rho_\tau^{\min}(\mathcal{A}/\sqrt{m})$ and $\rho_\tau^{\max}(\mathcal{A}/\sqrt{m})$ concentrate around 1 for relatively large m (but $\ll n_1 n_2$). More precisely, we have the following theorem:

Theorem 1 Let \mathcal{A} be an isotropic and subgaussian operator with some numerical constant L . Then there exists absolute constants c_1, c_2, c_3 depending on L only such that for any $\epsilon > 0$ and $m \geq 1$ satisfying

$$m \geq c_1 \frac{\tau n_2}{\epsilon^2}, \quad (9)$$

we have

$$\begin{aligned} 1 - c_2\epsilon &\leq \mathbb{E}\rho_\tau^{\min}\left(\frac{\mathcal{A}}{\sqrt{m}}\right) \\ &\leq \mathbb{E}\rho_\tau^{\max}\left(\frac{\mathcal{A}}{\sqrt{m}}\right) \leq 1 + c_2\epsilon \end{aligned} \quad (10)$$

and

$$\begin{aligned} \mathbb{P}\left[1 - \epsilon \leq \rho_\tau^{\min}\left(\frac{\mathcal{A}}{\sqrt{m}}\right) \leq \rho_\tau^{\max}\left(\frac{\mathcal{A}}{\sqrt{m}}\right) \leq 1 + \epsilon\right] \\ \geq 1 - \exp(-c_3\epsilon^2 m). \end{aligned} \quad (11)$$

Using the relation between the ℓ_* -constrained singular values and the mRIC, we have the following immediate corollary which was also established in [5] using different approaches :

Corollary 1 *Under the conditions of Theorem 1, there exists numerical constants c_1, c_2 such that for any $\epsilon > 0$ the mRIC constant $\delta_r(\mathcal{A}/\sqrt{m})$ satisfies*

$$\mathbb{P}\left[\delta_r(\mathcal{A}/\sqrt{m}) > \epsilon\right] \leq \exp(-c_1\epsilon^2 m), \quad (12)$$

as long as

$$m \geq c_2 \frac{rn_2}{\epsilon^2}. \quad (13)$$

Corollary 1 was established in [5] using an ϵ -net argument. The same procedure can not be generalized trivially to prove Theorem 1. The idea behind an ϵ -net argument is that any point in the set under consideration can be approximated using a point in its ϵ -cover with an error at most ϵ . One key ingredient in the proof given by [5] is that the approximation error matrix also has low rank. This is not satisfied by the error matrix in our case, because the difference between two matrices with small ℓ_* -ranks does not necessarily have a small ℓ_* -rank. This difficulty might be circumvented by resorting to the Dudley's inequality [22]. However, a good estimate of the covering number of the set $\{X \in \mathbb{R}^{n_1 \times n_2} : \|X\|_F = 1, \|X\|_*^2 \leq \tau\}$ that makes the Dudley's inequality tight enough is not readily available.

In Section V-A, we start with the Gaussian ensemble. We use the comparison theorems for Gaussian processes, the Gordon's inequality and the Slepian's inequality in particular, to show that the expected values of the ℓ_* -constrained singular values fall within the neighborhood of one. Then the concentration of measure phenomena in Gauss space immediately yields the

desired results of Theorem 1 for the Gaussian ensemble as both $\rho_\tau^{\min}(\cdot)$ and $\rho_\tau^{\max}(\cdot)$ are Lipschitz functions. The Gordon's and Slepian's inequalities rely heavily on Gaussian processes' highly symmetric property, which is not satisfied by general subgaussian processes.

The corresponding results for subgaussian operators are established in Section V-B. The problem is formulated as one of empirical processes. We then employ a recent general result of empirical processes established in [23]. The challenges of deriving Theorem 1 using more direct approaches such as Dudley's inequality and/or the general generic chaining bound [24] is also discussed.

One common ground of the proofs for the Gaussian case and the more general subgaussian case reveals the reason that \mathcal{A} is invertible on $\{\tau(X) \leq \tau\}$ while it is far from invertible on $\mathbb{R}^{n_1 \times n_2}$. Both proofs rely on the fact that the canonical Gaussian process indexed by the intersection of $\{\tau(X) \leq \tau\}$ and the unit sphere of $(\mathbb{R}^{n_1 \times n_2}, \|\cdot\|_F)$ can not go too far from zero in its life. This essentially means that the set $\{\tau(X) \leq \tau\} \cap \{\|X\|_F = 1\}$ with a small τ is significantly smaller than the unit sphere itself, on which the canonical Gaussian process would drift far away from zero. Refer to Section V for more precise meanings of these discussions.

B. Stability of Convex Relaxation Algorithms

In this section, we present the stability results for three convex relaxation algorithms: the matrix Basis Pursuit, the matrix Dantzig Selector, and the matrix LASSO estimator. These algorithms are reviewed briefly in Section III-C. As one will see in the proofs to Theorems 2, 3 and 4, the procedure of establishing these theorems has two steps:

- 1) Show that the error matrix $H = \hat{X} - X$ has a small ℓ_* -rank: $\tau(H) \leq \tau$ for some suitably selected τ , which automatically leads to a lower bound $\|\mathcal{A}(H)\|_2 \geq \rho_\tau \|H\|_F$. Here X is the true matrix and \hat{X} is its estimate given by convex relaxation algorithms.
- 2) Obtain an upper bound on $\|\mathcal{A}(H)\|_2$.

These are all relatively easy to show for the matrix Basis Pursuit algorithm. We have the following stability result:

Theorem 2 *If matrix X has rank r and the noise w is bounded; that is, $\|w\|_2 \leq \epsilon$, then the solution \hat{X} to the matrix Basis Pursuit (27) obeys*

$$\|\hat{X} - X\|_F \leq \frac{2\epsilon}{\rho_{8r}}. \quad (14)$$

The corresponding bound (36) using mRIC is expressed as $\frac{4\sqrt{1+\delta_{4r}}}{1-(1+\sqrt{2})\delta_{4r}} \cdot \varepsilon$ under the condition $\delta_{4r} \leq \sqrt{2} - 1$. Here δ_r is the mRIC defined in Definition 5. We note the ℓ_* -CMSV bound (14) is more concise and only requires $\rho_{8r} > 0$. Of course, we pay a price by replacing the subscript $4r$ with $8r$. A similar phenomena is also observed in the sparse signal reconstructions case [16]. As suggested by the numerical simulation in [16], by analogy we expect that it is easier to get $\rho_{8r} > 0$ than $\delta_{4r} \leq \sqrt{2} - 1$. However, we did not run simulations in this paper because it is not clear how to compute δ_{4r} within reasonable time even for small scale problems.

Before stating the results for the matrix Dantzig Selector and the matrix Lasso estimator, we cite a lemma of [5]:

Lemma 1 [5, Lemma 1.1] *Suppose $\mathbf{w} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_m)$. If $C \geq 4\sqrt{(1 + \rho_1^{\max}(\mathcal{A})) \log 12}$, then there exists a numerical constant $c > 0$ such that with probability greater than $1 - 2\exp(-cn_2)$ that*

$$\|\mathcal{A}^*(\mathbf{w})\| \leq C\sqrt{n_2}\sigma, \quad (15)$$

where \mathcal{A}^* is the adjoint operator of \mathcal{A} .

Lemma 1 allows to transform statements under the condition of $\|\mathcal{A}^*(\mathbf{w})\|_2 \leq \lambda$, e.g. Theorem 3 and 4, into ones that hold with large probability. We now present the error bounds for the matrix Dantzig Selector and the matrix LASSO estimator, whose proofs can be found in Sections IV-B and IV-C, respectively.

Theorem 3 *Suppose the noise vector in model (1) satisfies $\|\mathcal{A}^*(\mathbf{w})\|_2 \leq \lambda$, and suppose $X \in \mathbb{R}^{n_1 \times n_2}$ is of rank r . Then, the solution \hat{X} to the matrix Dantzig Selector (28) satisfies*

$$\|\hat{X} - X\|_F \leq \frac{4\sqrt{2}}{\rho_{8r}^2} \cdot \sqrt{r} \cdot \lambda. \quad (16)$$

Theorem 4 *Suppose the noise vector in model (1) satisfies $\|\mathcal{A}^*(\mathbf{w})\|_2 \leq \kappa\mu$ for some $\kappa \in (0, 1)$, and suppose $X \in \mathbb{R}^{n_1 \times n_2}$ is of rank r . Then, the solution \hat{X} to the matrix LASSO estimator (29) satisfies*

$$\|\hat{X} - X\|_F \leq \frac{1 + \kappa}{1 - \kappa} \cdot \frac{2\sqrt{2}}{\rho_{\frac{8r}{(1-\kappa)^2}}^2} \cdot \sqrt{r} \cdot \mu. \quad (17)$$

For example, if we take $\kappa = 1 - 2\sqrt{2}/3$, then the bound becomes

$$\|\hat{X} - X\|_F \leq 6(1 - \frac{\sqrt{2}}{3}) \cdot \frac{1}{\rho_{9r}} \cdot \sqrt{r} \cdot \mu. \quad (18)$$

The readers are encouraged to compare the statements of Theorem 3 and 4 with those using mRIC as cited in Section III-D (Equations (37), (38) and the conditions for them to be valid).

III. NOTATIONS, MEASUREMENT MODEL, RECONSTRUCTION ALGORITHMS, AND MATRIX RESTRICTED ISOMETRY CONSTANT

A. Notations

We use ℓ_p^m to denote the space \mathbb{R}^m equipped with the ℓ_p^m norm $\|\cdot\|_p$ defined as

$$\|\mathbf{x}\|_p = \left(\sum_{k \leq m} |x_k|^p \right)^{1/p} \quad \text{for } 1 \leq p < \infty \quad (19)$$

and

$$\|\mathbf{x}\|_\infty = \max_{k \leq m} |x_k| \quad (20)$$

for $\mathbf{x} = [x_1, \dots, x_m]^T \in \mathbb{R}^m$. The notation $\|\mathbf{x}\|_0$ counts the number of nonzero elements of \mathbf{x} .

Suppose $X = [\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_{n_2}] \in \mathbb{R}^{n_1 \times n_2}$. Define the Frobenius norm of X as $\|X\|_F = \sqrt{\sum_{i,j} |X_{ij}|^2} = \sqrt{\sum_i \sigma_i^2(X)}$, the nuclear norm as $\|X\|_* = \sum_i \sigma_i(X)$, and the operator norm as $\|X\|_2 = \max\{\sigma_i(X)\}$, where $\sigma_i(X)$ is the i th singular value of X . The rank of X is denoted by $\text{rank}(X) = \#\{i : \sigma_i(X) \neq 0\}$.

If we use $\boldsymbol{\sigma}(X) = [\sigma_1(X) \ \sigma_2(X) \ \dots \ \sigma_{n_1}(X)]^T$ to represent the singular value vector, then clearly we have the following relations:

$$\begin{aligned} \|X\|_F &= \|\boldsymbol{\sigma}(X)\|_2, \\ \|X\|_* &= \|\boldsymbol{\sigma}(X)\|_1, \\ \|X\|_2 &= \|\boldsymbol{\sigma}(X)\|_\infty, \\ \text{rank}(X) &= \|\boldsymbol{\sigma}(X)\|_0. \end{aligned} \quad (21)$$

Note that we use $\|\cdot\|_2$ to represent both the matrix operator norm and the ℓ_2 norm of a vector. The exact meaning can always be inferred from the context. These singular value vector

representations for the matrix norms immediately lead to

$$\begin{aligned}\|X\|_2 &\leq \|X\|_F \leq \|X\|_* \\ &\leq \sqrt{\text{rank}(X)} \|X\|_F \\ &\leq \text{rank}(X) \|X\|_2.\end{aligned}$$

The vectorization operator $\text{vec}(X) = [\mathbf{x}_1^T \ \mathbf{x}_2^T \ \dots \ \mathbf{x}_{n_2}^T]^T$ stacks the columns of X into a long column vector. The inner product of two matrices $X_1, X_2 \in \mathbb{R}^{n_1 \times n_2}$ is defined as

$$\begin{aligned}\langle X_1, X_2 \rangle &= \text{trace}(X_1^T X_2) \\ &= \text{vec}(X_1)^T \text{vec}(X_2) \\ &= \sum_{i,j} (X_1)_{ij} (X_2)_{ij}.\end{aligned}\tag{22}$$

The following Cauchy-Schwarz type inequalities are due to the fact that the dual norm of the Frobenius norm is the Frobenius norm, and the dual norm of the operator norm is the nuclear norm [3]:

$$\begin{aligned}\langle X_1, X_2 \rangle &\leq \|X_1\|_F \|X_2\|_F, \\ \langle X_1, X_2 \rangle &\leq \|X_1\|_* \|X_2\|_2.\end{aligned}$$

For any linear operator $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \mapsto \mathbb{R}^m$, its adjoint operator $\mathcal{A}^* : \mathbb{R}^m \mapsto \mathbb{R}^{n_1 \times n_2}$ is defined by the following relation

$$\langle \mathcal{A}(X), \mathbf{z} \rangle = \langle X, \mathcal{A}^*(\mathbf{z}) \rangle, \quad \forall X \in \mathbb{R}^{n_1 \times n_2}, \mathbf{z} \in \mathbb{R}^m.\tag{23}$$

A linear operator $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \mapsto \mathbb{R}^m$ can be represented by m matrices $\mathcal{A} = \{A_1, A_2, \dots, A_m\} \subset \mathbb{R}^{n_1 \times n_2}$ as follows

$$\mathcal{A}(X) = \begin{bmatrix} \langle A_1, X \rangle \\ \langle A_2, X \rangle \\ \vdots \\ \langle A_m, X \rangle \end{bmatrix}.\tag{24}$$

We will interchangeably use \mathcal{A} and \mathcal{A} to represent the same linear operator. The adjoint operation for this representation is given by

$$\mathcal{A}^*(\mathbf{z}) = \sum_{k=1}^m \mathbf{z}_k A_k \in \mathbb{R}^{n_1 \times n_2}.\tag{25}$$

In the derivations of this paper, we frequently use c, C, c_1, c_2, c_3 to represent numerical constants. For notation simplicity, the same symbol c could be used to represent different constants in the argument.

B. The Measurement Model

The following measurement model is used throughout the paper. Suppose we have a matrix $X \in \mathbb{R}^{n_1 \times n_2}$ with $\text{rank}(X) = r \ll n_1$. We observe $\mathbf{y} \in \mathbb{R}^m$ through the following linear model:

$$\mathbf{y} = \mathcal{A}(X) + \mathbf{w}, \quad (26)$$

where $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \mapsto \mathbb{R}^m$ is a linear operator and $\mathbf{w} \in \mathbb{R}^m$ is the noise/disturbance vector, either deterministic or random. In the deterministic setting we assume boundedness: $\|\mathbf{w}\|_2 \leq \varepsilon$, while in the stochastic setting we assume Gaussianity: $\mathbf{w} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_m)$. The model (26) is generally underdetermined with $m \ll n_1 n_2$, and the rank of X , r , is very small.

A fundamental problem pertaining to model (26) is to reconstruct the low-rank matrix X from the measurement \mathbf{y} by exploiting the low-rank property of X , and the stability of the reconstruction with respect to noise. For any reconstruction algorithm, we denote the estimate of X as \hat{X} , and the error matrix $H \stackrel{\text{def}}{=} \hat{X} - X$. In this paper, the stability problem aims to bound $\|H\|_F$ in terms of m, n_1, n_2, r , the linear operator \mathcal{A} , and the noise strength ε or σ^2 .

C. Reconstruction Algorithms

We briefly review three low-rank matrix recovery algorithms based on convex relaxation: the matrix Basis Pursuit, the matrix Dantzig selector, and the matrix LASSO estimator. A common theme of these algorithms is enforcing the low-rankness of solutions by penalizing large nuclear norms, or equivalently, the ℓ_1 norms of the singular value vectors. As a relaxation of the matrix rank, the nuclear norm remains a measure of low-rankness while being a convex function. In fact, the nuclear norm $\|\cdot\|_*$ is the convex envelop of $\text{rank}(\cdot)$ on the set $\{X \in \mathbb{R}^{n_1 \times n_2} : \|X\|_2 \leq 1\}$ [3, Theorem 2.2]. Most computational advantages of the aforementioned three algorithms result from the convexity of the nuclear norm. As demonstrated in Section IV, the nuclear norm enforcement guarantees low-rankness of the error matrix in the sense made precise in Section IV.

The matrix Basis Pursuit algorithm [3], [5] tries to minimize the nuclear norm of solutions subject to the measurement constraint. It is applicable to both noiseless settings and bounded

noise settings with a known noise bound ε . The matrix Basis Pursuit algorithm was originally developed for the noise-free case in [3], *i.e.*, $\varepsilon = 0$ in (27). In this paper, we refer to both cases as matrix Basis Pursuit. Mathematically, the matrix Basis Pursuit solves:

$$\text{mBP} : \min_{Z \in \mathbb{R}^{n_1 \times n_2}} \|Z\|_* \quad \text{subject to} \quad \|\mathbf{y} - \mathcal{A}(Z)\|_2 \leq \epsilon. \quad (27)$$

The matrix Dantzig selector [5] reconstructs a low-rank matrix when its linear measurements are corrupted by unbounded noise. Its estimate for X is the solution to the nuclear norm regularization problem:

$$\text{mDS} : \min_{Z \in \mathbb{R}^{n_1 \times n_2}} \|Z\|_* \quad \text{subject to} \quad \|\mathcal{A}^*(\mathbf{r})\|_2 \leq \lambda, \quad (28)$$

where $\mathbf{r} = \mathbf{y} - \mathcal{A}^*(\mathbf{z})$ is the residual vector, σ the noise standard deviation, and λ_n a control parameter.

The matrix LASSO estimator solves the following optimization problem [5], [25]:

$$\text{mLASSO} : \min_{Z \in \mathbb{R}^{n_1 \times n_2}} \frac{1}{2} \|\mathbf{y} - \mathcal{A}(Z)\|_2^2 + \lambda_n \|Z\|_*. \quad (29)$$

All three optimization problems can be solved using convex programs.

D. Matrix Restricted Isometry Constant

In this section, we present stability results related to matrix Basis Pursuit, the matrix Dantzig selector, and the matrix LASSO estimator. The aim of stability analysis is to derive error bounds of the solutions of these algorithms. These bounds usually involve the incoherence of the linear operator \mathcal{A} , which is measured by the mRIC defined below [3], [5]:

Definition 5 For each integer $r \in \{1, \dots, n_1\}$, the matrix restricted isometry constant (mRIC) δ_r of a linear operator $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \mapsto \mathbb{R}^m$ is defined as the smallest $\delta > 0$ such that

$$1 - \delta \leq \frac{\|\mathcal{A}(X)\|_2^2}{\|X\|_F^2} \leq 1 + \delta \quad (30)$$

holds for arbitrary non-zero matrix X of rank at most r .

A linear operator \mathcal{A} with a small δ_r roughly means that \mathcal{A} is nearly an isometry when restricted onto all matrices with rank at most r . Hence, it is no surprise that the mRIC is involved in the stability of recovering X from $\mathcal{A}(X)$ corrupted by noise when X is of rank at most r .

The Rayleigh quotient $\frac{\|\mathcal{A}(X)\|_2}{\|X\|_F}$ in the definition of the mRIC motivates us to define the rank constrained singular values, which are closely related to the mRIC.

Definition 6 For any integer $1 \leq r \leq n_1$ and linear operator $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \mapsto \mathbb{R}^m$, define the r -rank constrained minimal singular value ν_r^{\min} and r -rank constrained maximal singular value ν_r^{\max} of \mathcal{A} via

$$\nu_r^{\min}(\mathcal{A}) \stackrel{\text{def}}{=} \inf_{X \neq 0: \text{rank}(X) \leq r} \frac{\|\mathcal{A}(X)\|_2}{\|X\|_F}, \quad \text{and} \quad (31)$$

$$\nu_r^{\max}(\mathcal{A}) \stackrel{\text{def}}{=} \sup_{X \neq 0: \text{rank}(X) \leq r} \frac{\|\mathcal{A}(X)\|_2}{\|X\|_F}, \quad (32)$$

respectively.

The mRIC δ_r for linear operator \mathcal{A} is related to the r -rank constrained minimal and maximal singular values by

$$\delta_r = \max\{|1 - (\nu_r^{\min})^2|, |(\nu_r^{\max})^2 - 1|\}. \quad (33)$$

We comment that similar to the vector RIC, the exact computation of the mRIC is extremely difficult. Equation (3) implies that $\{X \neq 0 : \text{rank}(X) \leq r\} \subseteq \{X \neq 0 : \tau(X) \leq r\}$. As a consequence, the rank constrained singular values satisfy the following inequality

$$\rho_r^{\min} \leq \nu_r^{\min} \leq \nu_r^{\max} \leq \rho_r^{\max}, \quad (34)$$

which combined with (33) yields the following relationship between mRIC and ℓ_* -constrained singular values:

$$\delta_r \leq \max\{|1 - (\rho_r^{\min})^2|, |(\rho_r^{\max})^2 - 1|\}. \quad (35)$$

Now we cite stability results on the matrix Basis Pursuit, the matrix Dantzig selector, and the matrix LASSO estimator, which are expressed in terms of the mRIC. Assume X is of rank r and \hat{X} is its estimate given by any of the three algorithms; then we have the following:

- 1) matrix Basis Pursuit [5]: Suppose that $\delta_{4r} < \sqrt{2} - 1$ and $\|\mathbf{w}\|_2 \leq \varepsilon$. The solution to the matrix Basis Pursuit (27) satisfies

$$\|\hat{X} - X\|_F \leq \frac{4\sqrt{1 + \delta_{4r}}}{1 - (1 + \sqrt{2})\delta_{4r}} \cdot \varepsilon. \quad (36)$$

2) matrix Dantzig selector [5]: If $\delta_{4r} < \sqrt{2} - 1$ and $\|\mathcal{A}^*(\mathbf{w})\|_2 \leq \lambda$, then

$$\|\hat{X} - X\|_F \leq \frac{16}{1 - (\sqrt{2} + 1)\delta_{4r}} \cdot \sqrt{r} \cdot \lambda. \quad (37)$$

3) matrix LASSO estimator [5]: If $\delta_{4r} < (3\sqrt{2} - 1)/17$ and $\|\mathcal{A}^*(\mathbf{w})\| \leq \mu/2$, then the solution to the matrix LASSO (29) satisfies

$$\|\hat{X} - X\|_F \leq C(\delta_{4r})\sqrt{r} \cdot \mu, \quad (38)$$

for some numerical constant C .

Although the mRIC provides a measure quantifying the goodness of a linear operator, as mentioned earlier, its computation poses great challenges. In the literature, the computation issue is circumvented by resorting to a random argument. We cite one general result below [5]:

- Let $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \mapsto \mathbb{R}^m$ be a random linear operator satisfying the concentration inequality for any $X \in \mathbb{R}^{n_1 \times n_2}$ and $0 < \epsilon < 1$:

$$\mathbb{P}(|\|\mathcal{A}(X)\|_2^2 - \|X\|_F^2| \geq \epsilon\|X\|_F^2) \leq Ce^{-mc_0(\epsilon)}. \quad (39)$$

for fixed constant $C > 0$. Then, for any given $\delta \in (0, 1)$, there exist constants $c_1, c_2 > 0$ depending only on δ such that $\delta_r \leq \delta$, with probability not less than $1 - Ce^{-c_1 m}$, as long as

$$m \geq c_2 nr. \quad (40)$$

IV. STABILITY OF CONVEX RELAXATION BASED ON THE ℓ_1 -CONSTRAINED MINIMAL SINGULAR VALUE

In this section, we present the derivation of bounds on the reconstruction error for the matrix Basis Pursuit, the matrix Dantzig selector and the matrix LASSO estimator. As shown in Theorems 2, 3 and 4, our bounds are given in terms of the ℓ_* -CMSV rather than the mRIC of linear operator \mathcal{A} .

A. Basis Pursuit

In this section, we establish a bound on the Frobenius norm of matrix Basis Pursuit's error matrix using the ℓ_* -CMSV. Recall the two steps discussed in Section II-B:

- 1) Show that the error matrix $H = \hat{X} - X$ has small ℓ_* rank: $\tau(H) \leq 8r$, which automatically leads to a lower bound $\|\mathcal{A}(H)\|_2 \geq \rho_{8r}\|H\|_F$;
- 2) Obtain an upper bound on $\|\mathcal{A}(H)\|_2$.

For matrix Basis Pursuit (27), the second step is trivial as both X and \hat{X} satisfy constraint $\|\mathbf{y} - \mathcal{A}(Z)\| \leq \epsilon$ in (27). Therefore, the triangle inequality yields

$$\begin{aligned}
 \|\mathcal{A}(H)\|_2 &= \|\mathcal{A}(\hat{X} - X)\|_2 \\
 &\leq \|\mathcal{A}(\hat{X}) - \mathbf{y}\|_2 + \|\mathbf{y} - \mathcal{A}(X)\|_2 \\
 &\leq 2\epsilon.
 \end{aligned} \tag{41}$$

In order to establish that the error matrix has a small ℓ_* -rank in the first step, we present two lemmas on the properties of nuclear norms derived in [3]:

Lemma 2 [3, Lemma 2.3] *Let A and B be matrices of the same dimensions. If $AB^T = 0$ and $A^T B = 0$ then $\|A + B\|_* = \|A\|_* + \|B\|_*$.*

Lemma 3 [3, Lemma 3.4] *Let A and B be matrices of the same dimensions. Then there exist matrices B_1 and B_2 such that*

- 1) $B = B_1 + B_2$
- 2) $\text{rank}(B_1) \leq 2\text{rank}(A)$
- 3) $AB_2^T = 0$ and $A^T B_2 = 0$
- 4) $\langle B_1, B_2 \rangle = 0$.

Now we give a proof of Theorem 2:

Proof of Theorem 2: We decompose the error matrix $B = H$ according to Lemma 3 with $A = X$, more explicitly, we have:

- 1) $H = H_0 + H_c$
- 2) $\text{rank}(H_0) \leq 2\text{rank}(X) = 2r$
- 3) $XH_c^T = 0$ and $X^T H_c = 0$
- 4) $\langle H_0, H_c \rangle = 0$.

As observed by Recht *et.al* in [3] (See also [26], [5] and [16]), the fact that $\|\hat{X}\|_* = \|X + H\|_1$ is the minimum among all Z s satisfying the constraint in (27) implies that $\|H_c\|_*$ cannot be very

large. To see this, we observe that

$$\begin{aligned}
\|X\|_* &\geq \|X + H\|_* \\
&= \|X + H_c + H_0\|_* \\
&\geq \|X + H_c\|_* - \|H_0\|_* \\
&= \|X\|_* + \|H_c\|_* - \|H_0\|_*.
\end{aligned} \tag{42}$$

Here, for the last equality we used Lemma 2 and $XH_c^T = 0$, $X^T H_c = 0$. Therefore, we obtain

$$\|H_c\|_* \leq \|H_0\|_*, \tag{43}$$

which leads to

$$\begin{aligned}
\|H\|_* &\leq \|H_0\|_* + \|H_c\|_* \\
&\leq 2\|H_0\|_* \\
&\leq 2\sqrt{\text{rank}(H_0)}\|H_0\|_F \\
&= 2\sqrt{2r}\|H\|_F,
\end{aligned} \tag{44}$$

where for the next to the last inequality we used the fact that $\|H\|_* \leq \sqrt{\text{rank}(H)}\|H\|_F$, and for the last inequality we used the pythagoras theorem $\|H\|_F^2 = \|H_0\|_F^2 + \|H_c\|_F^2 \geq \|H_0\|_F^2$ because $\langle H_0, H_c \rangle = 0$. Inequality (44) is equivalent to

$$\tau(H) \leq 8 \text{rank}(X) = 8r. \tag{45}$$

It follows from (41) and Definition 2 that

$$\rho_{8r}\|H\|_F \leq \|\mathcal{A}(H)\|_2 \leq 2\epsilon. \tag{46}$$

Hence, we get the conclusion of Theorem 2

$$\|\hat{X} - X\|_F \leq \frac{2\epsilon}{\rho_{8r}}. \tag{47}$$

■

B. Dantzig Selector

This subsection is devoted to the proof of Theorem 3:

Proof of Theorem 3: Suppose $X \in \mathbb{R}^{n_1 \times n_2}$ is of rank r , and \hat{X} is the solution to the matrix Dantzig selector (28). Define $H = \hat{X} - X$. We note that to obtain that H has a small ℓ_* -rank (45), we used only two conditions:

- $\|\hat{X}\|_* = \|X + H\|_*$ is the minimum among all matrices satisfying the optimization constraint;
- the true signal X satisfies the constraint.

Obviously, the first condition holds simply because of the structure of the matrix Dantzig selector. If the noise vector \mathbf{w} satisfies $\|\mathcal{A}(\mathbf{w})\|_2 \leq \lambda$, then the true signal X also satisfy the constraint:

$$\begin{aligned} \|\mathcal{A}^*(\mathbf{r})\|_2 &= \|\mathcal{A}^*(\mathbf{y} - \mathcal{A}(X))\|_2 \\ &= \|\mathcal{A}^*(\mathbf{w})\|_2 \leq \lambda. \end{aligned} \quad (48)$$

Consequently, we have $\tau(H) \leq 8r$ following the same procedure as in the Proof of Theorem 2 in Section IV-A, or equivalently,

$$\|H\|_* \leq \sqrt{8r} \|H\|_F. \quad (49)$$

We now turn to the second step to obtain an upper bound on $\|\mathcal{A}(X)\|_F$. The condition $\|\mathcal{A}(\mathbf{w})\|_2 \leq \lambda$ and the constraint in the Dantzig selector (28) yield

$$\|\mathcal{A}^*(\mathcal{A}(H))\|_2 \leq 2\lambda \quad (50)$$

because

$$\begin{aligned} \mathcal{A}^*(\mathbf{w} - \hat{\mathbf{r}}) &= \mathcal{A}^*\left((\mathbf{y} - \mathcal{A}(X)) - (\mathbf{y} - \mathcal{A}(\hat{X}))\right) \\ &= \mathcal{A}^*\left(\mathcal{A}(\hat{X}) - \mathcal{A}(X)\right) = \mathcal{A}^*(\mathcal{A}(H)), \end{aligned} \quad (51)$$

where $\hat{\mathbf{r}} = \mathbf{y} - \mathcal{A}(\hat{X})$ is the residual corresponding to the matrix Dantzig selector solution \hat{X} . Therefore, we obtain an upper bound on $\|\mathcal{A}(H)\|_F^2$ as follows:

$$\begin{aligned} \langle \mathcal{A}(H), \mathcal{A}(H) \rangle &= \langle H, \mathcal{A}^*(\mathcal{A}(H)) \rangle \\ &\leq \|H\|_* \|\mathcal{A}^*(\mathcal{A}(H))\|_2 \\ &\leq 2\lambda \|H\|_*. \end{aligned} \quad (52)$$

Equation (52), the definition of ρ_{8r} , and equation (49) together yield

$$\begin{aligned}\rho_{8r}^2 \|H\|_F^2 &\leq \langle \mathcal{A}(H), \mathcal{A}(H) \rangle \\ &\leq 2\lambda \|H\|_* \\ &\leq 2\lambda \sqrt{8r} \sigma \|H\|_F.\end{aligned}\tag{53}$$

We conclude that

$$\|H\|_F \leq \frac{4\sqrt{2}}{\rho_{8r}^2} \cdot \sqrt{r} \cdot \lambda,\tag{54}$$

which is exactly the result of Theorem 3. ■

C. LASSO Estimator

We derive a bound on the matrix LASSO estimator using the procedure developed in [5] (see also [27]).

Proof of Theorem 4: Suppose the noise \mathbf{w} satisfies $\|\mathcal{A}^*(\mathbf{w})\|_2 \leq \kappa\mu$ for some small $\kappa > 0$. Because \hat{X} is a solution to (29), we have

$$\frac{1}{2} \|\mathcal{A}(\hat{X}) - \mathbf{y}\|_2^2 + \mu \|\hat{X}\|_* \leq \frac{1}{2} \|\mathcal{A}(X) - \mathbf{y}\|_2^2 + \mu \|X\|_*.$$

Consequently, substituting $\mathbf{y} = \mathcal{A}(X) + \mathbf{w}$ yields

$$\begin{aligned}\mu \|\hat{X}\|_* &\leq \frac{1}{2} \|\mathcal{A}(X) - \mathbf{y}\|_2^2 - \frac{1}{2} \|\mathcal{A}(\hat{X}) - \mathbf{y}\|_2^2 + \mu \|X\|_* \\ &= \frac{1}{2} \|\mathbf{w}\|_2^2 - \frac{1}{2} \|\mathcal{A}(\hat{X} - X) - \mathbf{w}\|_2^2 + \mu \|X\|_* \\ &= \frac{1}{2} \|\mathbf{w}\|_2^2 - \frac{1}{2} \|\mathcal{A}(\hat{X} - X)\|_2^2 \\ &\quad + \langle \mathcal{A}(\hat{X} - X), \mathbf{w} \rangle - \frac{1}{2} \|\mathbf{w}\|_2^2 + \mu \|X\|_* \\ &\leq \langle \mathcal{A}(\hat{X} - X), \mathbf{w} \rangle + \mu \|X\|_* \\ &= \langle \hat{X} - X, \mathcal{A}^*(\mathbf{w}) \rangle + \mu \|X\|_*.\end{aligned}$$

Using the Cauchy-Swcharz type inequality, we get

$$\begin{aligned}\mu \|\hat{X}\|_* &\leq \|\hat{X} - X\|_* \|\mathcal{A}^*(\mathbf{w})\|_2 + \mu \|X\|_* \\ &= \kappa\mu \|H\|_* + \mu \|X\|_*,\end{aligned}$$

which leads to

$$\|\hat{X}\|_* \leq \kappa \|H\|_* + \|X\|_*.$$

Therefore, similar to the argument in (42) we have

$$\begin{aligned} \|X\|_* &\geq \|\hat{X}\|_* - \kappa \|H\|_* \\ &= \|X + H\|_* - \kappa \|H\|_* \\ &\geq \|X + H_c + H_0\|_* - \kappa (\|H_c\|_* + \|H_0\|_*) \\ &\geq \|X + H_c\|_* - \|H_0\|_* - \kappa (\|H_c\|_* + \|H_0\|_*) \\ &= \|X\|_* + \|H_c\|_* - \|H_0\|_* - \kappa (\|H_c\|_* + \|H_0\|_*) \\ &= \|X\|_* + (1 - \kappa)\|H_c\|_* - (1 + \kappa)\|H_0\|_*. \end{aligned}$$

Consequently, we have

$$\|H_c\|_* \leq \frac{1 + \kappa}{1 - \kappa} \|H_0\|_*,$$

an inequality slightly worse than (43) for small κ . Therefore, an argument similar to the one leading to (44) yields

$$\|H\|_* \leq \frac{2}{1 - \kappa} \sqrt{2r} \|H\|_F, \quad (55)$$

or equivalently,

$$\tau(H) \leq \frac{8r}{(1 - \kappa)^2}. \quad (56)$$

Now we need to establish a bound on

$$\begin{aligned} \|\mathcal{A}^*(\mathcal{A}(H))\|_2 &\leq \|\mathcal{A}^*(\mathbf{y} - \mathcal{A}(X))\|_2 + \|\mathcal{A}^*(\mathbf{y} - \mathcal{A}(\hat{X}))\|_2 \\ &\leq \|\mathcal{A}^*(\mathbf{w})\|_2 + \|\mathcal{A}^*(\mathbf{y} - \mathcal{A}(\hat{X}))\|_2 \\ &= \kappa\mu + \|\mathcal{A}^*(\mathbf{y} - \mathcal{A}(\hat{X}))\|_2. \end{aligned} \quad (57)$$

We follow the procedure in [5] (see also [27]) to estimate $\|\mathcal{A}^*(\mathbf{y} - \mathcal{A}(\hat{X}))\|_2$. Since \hat{X} is the solution to (29), the optimality condition yields that

$$\mathcal{A}^*(\mathbf{y} - \mathcal{A}(\hat{X})) \in \partial\|\hat{X}\|_*, \quad (58)$$

where $\partial\|\hat{X}\|_*$ is the family of subgradient of $\|\cdot\|_*$ evaluated at \hat{X} . According to [10], if the singular value decomposition of \hat{X} is $U\Sigma V^T$, then we have

$$\begin{aligned}\partial\|\hat{X}\|_* &= \{\mu(UV^T + W) : \|W\|_2 \leq 1, \\ &\quad U^T W = 0, W V = 0\}.\end{aligned}\tag{59}$$

As a consequence, we obtain $\mathcal{A}^*(\mathbf{y} - \mathcal{A}(\hat{X})) = \mu(UV^T + W)$ and

$$\begin{aligned}\|\mathcal{A}^*(\mathbf{y} - \mathcal{A}(\hat{X}))\|_2 &\leq \|\mu(UV^T + W)\|_2 \\ &= \mu.\end{aligned}\tag{60}$$

We used $\|UV^T + W\|_2 = 1$ because

$$\begin{aligned}&\max_{\mathbf{x}: \|\mathbf{x}\|_2=1} \|(UV^T + W)\mathbf{x}\|_2 \\ &= \max_{\mathbf{y}: \|\mathbf{y}\|_2=1} \|(UV^T + W)V\mathbf{y}\|_2 \leq 1.\end{aligned}\tag{61}$$

Following the same lines in (52), we get

$$\|\mathcal{A}(H)\|_2^2 \leq (\kappa + 1)\mu\|H\|_*.\tag{62}$$

Then, Equation (55), (57) and (60)

$$\begin{aligned}\rho_{\frac{8r}{(1-\kappa)^2}}^2 \|H\|_F^2 &\leq \|\mathcal{A}(H)\|_2^2 \\ &\leq (\kappa + 1)\mu \frac{\sqrt{8r}}{1 - \kappa} \|H\|_F.\end{aligned}\tag{63}$$

As a consequence, the conclusion of Theorem 4 holds. ■

V. ℓ_1 -CONSTRAINED SINGULAR VALUES OF RANDOM MEASUREMENT ENSEMBLES

This section is devoted to analyzing the properties of the ℓ_* -CMSVs for several important random sensing ensembles. Although the bounds in Theorem 2, 3 and 4 have concise forms, they are useless if the quantity involved, ρ_τ , is zero or approaches zero for most matrices as n_1, n_2, m, k vary in a reasonable manner. We show that, at least for the isotropic and subgaussian ensemble, the ℓ_* -CMSVs are bounded away from zero with high probability.

A. Gaussian sensing matrix

We start with the Gaussian ensembles. The derivation mimics the one for estimating maximal and minimal singular values for rectangular random matrix with entries from Gaussian distribution. The argument is due to Gordon and can be found in [28] and [29]. We make modifications to take the ℓ_* constraint into account. Due to the highly symmetric properties of Gaussian random variables, we expect simplified derivations for the Gaussian ensemble.

Consider a linear operator

$$\begin{aligned} \mathcal{A} : \quad (\mathbb{R}^{n_1 \times n_2}, \|\cdot\|_*) &\rightarrow \ell_2^m \\ X &\mapsto \mathcal{A}(X) \end{aligned} \quad (64)$$

where \mathcal{A} is represented by a collection of m matrices $\mathcal{A} = \{A_1, A_2, \dots, A_m\} \subset \mathbb{R}^{n_1 \times n_2}$. We assume that the entries of matrices $A_k \in \mathbb{R}^{n_1 \times n_2}$ are *i.i.d.* Gaussian random variables from $\mathcal{N}(0, 1)$. Our goal is to study the behavior of the ℓ_* -CMSV for \mathcal{A}/\sqrt{m} . We say that the operator \mathcal{A}/\sqrt{m} comes from the *Gaussian ensemble*.

We note that the ℓ_* -constrained minimal and maximal singular values can be equivalently expressed as

$$\rho_\tau^{\min}(\mathcal{A}) = \inf_{X \in \mathcal{H}_\tau} \sup_{v \in S^{m-1}} \langle \mathcal{A}(X), v \rangle, \quad \text{and} \quad (65)$$

$$\rho_\tau^{\max}(\mathcal{A}) = \sup_{X \in \mathcal{H}_\tau} \sup_{v \in S^{m-1}} \langle \mathcal{A}(X), v \rangle. \quad (66)$$

Here the set

$$\mathcal{H}_\tau \stackrel{\text{def}}{=} \{X \in \mathbb{R}^{n_1 \times n_2} : \|X\|_F = 1, \|X\|_*^2 \leq \tau\} \quad (67)$$

is a subset of the unit sphere of $(\mathbb{R}^{n_1 \times n_2}, \|\cdot\|_F)$, and S^{m-1} is the unit sphere in ℓ_2^m .

Clearly, the quantity $\langle \mathcal{A}(X), v \rangle$ is a Gaussian random variable for fixed $X \in \mathcal{H}_\tau$ and $v \in S^{m-1}$. As a consequence, the problem of estimating $\rho_\tau^{\min}(\mathcal{A})$ and $\rho_\tau^{\max}(\mathcal{A})$ reduces to the study of a Gaussian process $\xi_{X,v}$ indexed by $\mathcal{H}_\tau \times S^{m-1}$, especially the behavior of its extremal values. Suitable tools for this purpose are Gaussian comparison theorems, in particular the Gordon's inequality and the Slepian's inequality. Under certain conditions on the second order moments of the increments, these two inequalities compare the expected extremal values of a Gaussian process with another Gaussian process, one that is simpler for analysis purposes.

Lemma 4 [22, Chapter 3.1] Suppose $(\xi_{u,v})_{u \in U, v \in V}$ and $(\zeta_{u,v})_{u \in U, v \in V}$ be Gaussian processes with zero mean. If for all $u, u' \in U$ and $v, v' \in V$,

$$\mathbb{E}(\xi_{u,v} - \xi_{u',v'})^2 \leq \mathbb{E}(\zeta_{u,v} - \zeta_{u',v'})^2, \quad (68)$$

then

$$\text{Slepian's inequality: } \mathbb{E} \sup_{u \in U, v \in V} \xi_{u,v} \leq \mathbb{E} \sup_{u \in U, v \in V} \zeta_{u,v}. \quad (69)$$

If

$$\begin{aligned} \mathbb{E}(\xi_{u,v} - \xi_{u',v'})^2 &\leq \mathbb{E}(\zeta_{u,v} - \zeta_{u',v'})^2, \quad \text{if } u \neq u' \\ \mathbb{E}(\xi_{u,v} - \xi_{u,v'})^2 &= \mathbb{E}(\zeta_{u,v} - \zeta_{u,v'})^2, \end{aligned} \quad (70)$$

then

$$\text{Gordon's inequality: } \mathbb{E} \inf_{u \in U} \sup_{v \in V} X_{u,v} \geq \mathbb{E} \inf_{u \in U} \sup_{v \in V} Y_{u,v}. \quad (71)$$

Consider only one index set and assume the index is time. The Slepian's inequality states that, the Gaussian process with the bigger step size measured by the second order moments of the increments will have a larger maximal distance away from the origin in its life. The Gordon's inequality can be understood in a similar manner.

To the end of analyzing $\rho_\tau^{\min}(\mathcal{A})$ and $\rho_\tau^{\max}(\mathcal{A})$, define two Gaussian processes

$$\begin{aligned} \xi_{X,v} &= \langle \mathcal{A}(X), \mathbf{v} \rangle \\ &= \sum_{\substack{i \leq n_1 \\ j \leq n_2}} \sum_{k \leq m} (A_k)_{ij} X_{ij} \mathbf{v}_k, \quad X \in \mathcal{H}_\tau, \mathbf{v} \in S^{m-1}, \end{aligned} \quad (72)$$

$$\begin{aligned} \zeta_{X,v} &= \langle G, X \rangle + \langle \mathbf{h}, \mathbf{v} \rangle \\ &= \sum_{\substack{i \leq n_1 \\ j \leq n_2}} G_{ij} X_{ij} + \sum_{k \leq m} \mathbf{h}_k \mathbf{v}_k, \quad X \in \mathcal{H}_\tau, \mathbf{v} \in S^{m-1}, \end{aligned} \quad (73)$$

where \mathcal{A} is a Gaussian linear operator, and G is a Gaussian matrix and \mathbf{h} is a Gaussian vector.

Clearly, $\mathbb{E}\xi_{u,v} = \mathbb{E}\zeta_{u,v} = 0$. We will focus on the application of Gordon's lemma to study ρ_τ^{\min} . The analysis of ρ_τ^{\max} using Slepian's lemma follows a similar and actually simpler argument.

Algebraic manipulations using the fact $\|X\|_F = 1$, $\|\mathbf{v}\|_2 = 1$ and Hölder's inequality show that

$$\begin{aligned} \mathbb{E}(\xi_{X,\mathbf{v}} - \xi_{X',\mathbf{v}'})^2 &= \sum_{i,j,k} (X_{ij}\mathbf{v}_k - X'_{ij}\mathbf{v}'_k)^2 \\ &\leq \sum_{i,j} (X_{ij} - X'_{ij})^2 + \sum_k (\mathbf{v}_k - \mathbf{v}'_k)^2 \\ &= \mathbb{E}(\zeta_{X,\mathbf{v}} - \zeta_{X',\mathbf{v}'})^2, \end{aligned} \quad (74)$$

Furthermore, setting $X = X'$ immediately gives

$$\begin{aligned} \mathbb{E}(\xi_{X,\mathbf{v}} - \xi_{X,\mathbf{v}'})^2 &= \sum_k (\mathbf{v}_k - \mathbf{v}'_k)^2 \\ &= \mathbb{E}(\zeta_{X,\mathbf{v}} - \zeta_{X,\mathbf{v}'})^2. \end{aligned} \quad (75)$$

As a consequence, Gordon's inequality (71) implies that

$$\begin{aligned} \mathbb{E}\rho_\tau^{\min}(\mathcal{A}) &= \mathbb{E} \inf_{X \in \mathcal{H}_\tau} \sup_{\mathbf{v} \in S^{m-1}} \xi_{X,\mathbf{v}} \\ &\geq \mathbb{E} \inf_{X \in \mathcal{H}_\tau} \sup_{\mathbf{v} \in S^{m-1}} \zeta_{X,\mathbf{v}}. \end{aligned} \quad (76)$$

To compute the right hand side of the previous formula, we proceed as

$$\begin{aligned} &\mathbb{E} \inf_{X \in \mathcal{H}_\tau} \sup_{\mathbf{v} \in S^{m-1}} \zeta_{X,\mathbf{v}} \\ &= \mathbb{E} \inf_{X \in \mathcal{H}_\tau} \langle G, X \rangle + \mathbb{E} \sup_{\mathbf{v} \in S^{m-1}} \langle \mathbf{h}, \mathbf{v} \rangle \\ &\geq -\mathbb{E} \sup_{X \in \mathcal{H}_\tau} \|G\|_2 \|X\|_* + \mathbb{E} \|\mathbf{h}\|_2 \\ &= -\sqrt{\tau} \mathbb{E} \|G\|_2 + \mathbb{E} \|\mathbf{h}\|_2 \\ &\geq -3\sqrt{\tau} \sqrt{n_2} + \sqrt{m}. \end{aligned} \quad (77)$$

We need some extra effort for the last inequality. Recall the convention $n_1 \leq n_2$ and note that

$$\mathbb{E} \|G\|_2 \leq \sqrt{n_1} + \sqrt{n_2} \leq 2\sqrt{n_2}, \quad (78)$$

$$\mathbb{E} \|\mathbf{h}\|_2 = \sqrt{2} \frac{\Gamma(\frac{m+1}{2})}{\Gamma(\frac{m}{2})} \begin{cases} \geq \sqrt{\frac{2}{\pi}} \sqrt{m}; \\ \leq \sqrt{m}, \end{cases} \quad (79)$$

where for (78) we used an upper bound for the expected largest singular value of a rectangular Gaussian matrix [28], [29], and equation (79) comes from an explicit computation involving χ^2 distribution. It suffices to show that

$$3\sqrt{n_2} - \mathbb{E} \|G\|_2 \geq \mathbb{E} \|\mathbf{h}\|_2 - \sqrt{m}. \quad (80)$$

To this end, define $\mathbf{g} \sim \mathcal{N}(0, \mathbf{I}_{n_2})$ and assume $m \geq n_2$. Since $\sqrt{m} - \sqrt{2}\Gamma((m+1)/2)/\Gamma(m/2)$ is decreasing in m , we have

$$\begin{aligned} \sqrt{m} - \mathbb{E}\|\mathbf{h}\|_2 &\leq \sqrt{n_2} - \mathbb{E}\|\mathbf{g}\|_2 \\ &\leq \left(1 - \sqrt{\frac{2}{\pi}}\right) \sqrt{n_2} \\ &\leq 3\sqrt{n_2} - 2\sqrt{n_2} \\ &\leq 3\sqrt{n_2} - \mathbb{E}\|G\|_2. \end{aligned} \tag{81}$$

Consequently, inequality (77) holds under that condition that $m \geq n_2$.

Therefore, Gordon's lemma yields for the numerical constant $c = 3$

$$\mathbb{E}\rho_\tau^{\min}\left(\frac{\mathcal{A}}{\sqrt{m}}\right) = \frac{1}{\sqrt{m}}\mathbb{E}\inf_{X \in \mathcal{H}_\tau} \|\mathcal{A}(X)\|_2 \geq 1 - c\sqrt{\frac{\tau n_2}{m}}, \tag{82}$$

as long as $m \geq n_2$. Note that the bound (82) is non-trivial only if $m \geq c^2 \tau n_2$ since otherwise we can use 0 as a lower bound. So we can drop the condition $m \geq n_2$.

Similarly, using Slepian's lemma, we conclude a similar bound for the largest singular value:

$$\mathbb{E}\rho_\tau^{\max}\left(\frac{\mathcal{A}}{\sqrt{m}}\right) = \frac{1}{\sqrt{m}}\mathbb{E}\sup_{X \in \mathcal{H}_\tau} \|\mathcal{A}(X)\|_2 \leq 1 + c\sqrt{\frac{\tau n_2}{m}}. \tag{83}$$

Now we have shown that $\rho_\tau^{\min}(\mathcal{A}/\sqrt{m})$ and $\rho_\tau^{\max}(\mathcal{A}/\sqrt{m})$ are close to one in average. What we really want is to show that the typical values of $\rho_\tau^{\min}(\mathcal{A}/\sqrt{m})$ and $\rho_\tau^{\max}(\mathcal{A}/\sqrt{m})$, rather than their averages, are close to one. Put in another way, we need to exclude the possibility of the "lottery phenomenon" [30] where an event with small probability contributes most of the average. To this end, we show that the largest and smallest constrained singular values of \mathcal{A} are 1-Lipschitz functions. By the concentration of measure phenomena, nice functions (e.g. Lipschitz functions) that depend on many parameters are almost constant. As a consequence, ρ_τ^{\min} and ρ_τ^{\max} would concentrate their values around their means.

Note a function F defined on a metric space (\mathcal{X}, d) is called *L-Lipschitz* if $|F(x) - F(x')| \leq Ld(x, x')$ holds for all $x, x' \in X$. We have the following lemma:

Lemma 5 *The ℓ_* -constrained singular values $\rho_\tau^{\max}(\cdot)$ and $\rho_\tau^{\min}(\cdot)$ are 1-Lipschitz in $(\mathbb{R}^{m \times n_1 \times n_2}, d)$*

when the metric $d(\mathcal{A}, \mathcal{B})$ is defined as

$$d(\mathcal{A}, \mathcal{B}) = \left(\sum_{k=1}^m \|A_k - B_k\|_F^2 \right)^{1/2}. \quad (84)$$

More precisely, we have

$$\begin{aligned} |\rho_\tau^{\max}(\mathcal{A}) - \rho_\tau^{\max}(\mathcal{B})| &\leq d(\mathcal{A}, \mathcal{B}) \\ |\rho_\tau^{\min}(\mathcal{A}) - \rho_\tau^{\min}(\mathcal{B})| &\leq d(\mathcal{A}, \mathcal{B}). \end{aligned} \quad (85)$$

Proof: Note that

$$\begin{aligned} \rho_\tau^{\max}(\mathcal{A}) &= \sup_{X \in \mathcal{H}_\tau} \|\mathcal{A}(X)\|_2 \\ &= \sup_{X \in \mathcal{H}_\tau} \|\mathcal{B}(X) + (\mathcal{A} - \mathcal{B})(X)\|_2 \\ &\leq \sup_{X \in \mathcal{H}_\tau} \|\mathcal{B}(X)\| + \sup_{X \in \mathcal{H}_\tau} \|(\mathcal{A} - \mathcal{B})(X)\|_2 \\ &\leq \rho_\tau^{\max}(\mathcal{B}) + \sup_{X \in \mathcal{H}_\tau} \left(\sum_{k=1}^m |\langle (A_k - B_k), X \rangle|^2 \right)^{1/2} \\ &\leq \rho_\tau^{\max}(\mathcal{B}) + \sup_{X \in \mathcal{H}_\tau} \left(\sum_{k=1}^m \|A_k - B_k\|_F^2 \|X\|_F^2 \right)^{1/2} \\ &= \rho_\tau^{\max}(\mathcal{B}) + d(\mathcal{A}, \mathcal{B}). \end{aligned} \quad (86)$$

Due to the symmetric role of \mathcal{A} and \mathcal{B} , we conclude that

$$|\rho_\tau^{\max}(\mathcal{A}) - \rho_\tau^{\max}(\mathcal{B})| \leq d(\mathcal{A}, \mathcal{B}). \quad (87)$$

Using similar arguments, we establish

$$|\rho_\tau^{\min}(\mathcal{A}) - \rho_\tau^{\min}(\mathcal{B})| \leq d(\mathcal{A}, \mathcal{B}). \quad (88)$$

■

By the functional form of concentration of measure in Gauss space, we have

$$\begin{aligned} &\mathbb{P} [\rho_\tau^{\max}(\mathcal{A}) - \mathbb{E} \rho_\tau^{\max}(\mathcal{A}) > t] \\ &= \mathbb{P} [\rho_\tau^{\max}(\mathcal{A}) > \sqrt{m} + c\sqrt{\tau n_2} + t] \\ &\leq \exp(-t^2/2), \end{aligned} \quad (89)$$

and

$$\begin{aligned}
& \mathbb{P} [\rho_{\tau}^{\min}(\mathcal{A}) - \mathbb{E}\rho_{\tau}^{\min}(\mathcal{A}) < -t] \\
&= \mathbb{P} [\rho_{\tau}^{\min}(\mathcal{A}) < \sqrt{m} - c\sqrt{\tau n_2} - t] \\
&\leq \exp(-t^2/2).
\end{aligned} \tag{90}$$

Therefore, replacing t with \sqrt{mt} we have

$$\begin{aligned}
& \mathbb{P} \left[\rho_{\tau}^{\max}(\mathcal{A}/\sqrt{m}) > 1 + c\sqrt{\frac{\tau n_2}{m}} + t \right] \leq \exp(-mt^2/2), \\
& \mathbb{P} \left[\rho_{\tau}^{\min}(\mathcal{A}/\sqrt{m}) < 1 - c\sqrt{\frac{\tau n_2}{m}} - t \right] \leq \exp(-mt^2/2).
\end{aligned} \tag{91}$$

If we set

$$\begin{aligned}
m &\geq 4c^2 \frac{\tau n_2}{\epsilon^2} \\
t &= \frac{\epsilon}{2},
\end{aligned} \tag{92}$$

with $\epsilon \in (0, 1)$ we have

$$\begin{aligned}
& \mathbb{P} [\rho_{\tau}^{\max}(\mathcal{A}/\sqrt{m}) > 1 + \epsilon] \leq \exp(-m\epsilon^2/8), \\
& \mathbb{P} [\rho_{\tau}^{\min}(\mathcal{A}/\sqrt{m}) < 1 - \epsilon] \leq \exp(-m\epsilon^2/8).
\end{aligned} \tag{93}$$

Therefore, for the Gaussian ensemble, the ℓ_* -constrained singular values of \mathcal{A}/\sqrt{m} concentrate around 1 with large probability.

B. Sub-Gaussian sensing matrix

In this section, we establish results for isotropic and subgaussian measurement operators that are comparable to those for the Gaussian ensemble presented in the previous section. The techniques used in Section V-A, for example, the Slepian's and Gordon's inequalities, and the measure of concentration phenomenon in Gauss space, depend heavily on the highly symmetric property of Gaussian random variables and can not be extended to the subgaussian case trivially. In this section, we employ a recent estimate on the behavior of empirical processes involving subgaussian random variables. Before going into details of this result established in [23], we first formulate the ℓ_* -CMSV for subgaussian measurement operators and discuss the difficulties.

Suppose the linear operator \mathcal{A} is generated in a way such that $\mathbb{E}\|\mathcal{A}(X)\|_2^2 = m\|X\|_F^2$ for any $X \in \mathbb{R}^{n_1 \times n_2}$. We note that $\rho_\tau^{\max}(\mathcal{A}/\sqrt{m}) \leq 1 + \epsilon$ and $\rho_\tau^{\min}(\mathcal{A}/\sqrt{m}) > 1 - \epsilon$ are equivalent with

$$\begin{aligned} & \sup_{X \in \mathcal{H}_\tau} \left| \frac{1}{m} \mathcal{A}(X)^T \mathcal{A}(X) - 1 \right| \\ &= \sup_{X \in \mathcal{H}_\tau} \left| \frac{1}{m} \sum_{k=1}^m \langle A_k, X \rangle^2 - 1 \right| \leq \epsilon. \end{aligned} \quad (94)$$

As usual, the operator \mathcal{A} is represented by a collection of matrices $\mathcal{A} = \{A_1, \dots, A_m\}$. We define a class of functions parameterized by X as $\mathcal{F}_\tau \stackrel{\text{def}}{=} \{f_X(\cdot) = \langle X, \cdot \rangle : X \in \mathcal{H}_\tau\}$. Recall that

$$\mathcal{H}_\tau = \{X \in \mathbb{R}^{n_1 \times n_2} : \|X\|_F = 1, \|X\|_*^2 \leq \tau\}. \quad (95)$$

Denote P_m the empirical measure that puts equal mass at each of the m random observations A_1, \dots, A_m , *i.e.*,

$$P_m(\cdot) = \frac{1}{m} \sum_{i=1}^m \delta_{A_i}(\cdot) \quad (96)$$

with $\delta_A(\cdot)$ the dirac measure on $\mathbb{R}^{n_1 \times n_2}$ that puts unit mass at A . We realize that $\{\frac{1}{m} \sum_{i=1}^m \langle A_i, X \rangle^2\}$ is nothing but the empirical process $\{P_m(f^2)\}_{f \in \mathcal{F}_\tau}$. We slightly abuse notation and use $\mathbb{E}f^2$ to denote $\mathbb{E}f^2(A)$. Then, our goal is to estimate

$$\mathbb{E} \sup_{f \in \mathcal{F}_\tau} |P_m(f^2) - \mathbb{E}f^2| \quad (97)$$

and

$$\mathbb{P} \left\{ \sup_{f \in \mathcal{F}_\tau} |P_m(f^2) - \mathbb{E}f^2| \right\}, \quad (98)$$

a central topic of the study of empirical processes.

Another slightly different (but closely related) way to view (94) is to consider

$$\xi_X \stackrel{\text{def}}{=} \left| \frac{1}{m} \sum_{i=1}^m \langle A_i, X \rangle^2 - 1 \right|, X \in \mathcal{H}_\tau \quad (99)$$

as a random process. Then $\mathbb{E}\xi_X$ can be computed using, for example, Dudley's inequality [22], [24], which involves the entropy number or covering number of \mathcal{H}_τ [31], [32]. This idea is used to establish bounds on the minimal and maximal singular values for rectangular random matrices. There are several challenges associated with this approach for estimating the ℓ_* -CMSV of general isotropic and subgaussian operators. First, a good estimate of the entropy number of \mathcal{H}_τ is not readily available. Second, even if the elements of the sensing operator are

subgaussian random variables, the increments $|\xi_X - \xi_Z|$ are actually a mixture of subexponential and subgaussian random variables, instead of being pure subgaussian. This causes problem when the Dudley's inequality, or the better generic chaining bound presented later, is applied. In addition, a corresponding concentration of measure result for subgaussian distributions is also absent.

Before we turn to the general empirical process result of [23] developed by delicate use of the powerful generic chaining idea, we need some notations and definitions. A key concept in studying general Gaussian processes as well as the empirical process $\{P_m(f^2)\}_{f \in \mathcal{F}_\tau}$ is the γ_p functional we are going to define. We need some setup first. For any set \mathcal{X} , an *admissible sequence* is a sequence of increasing partitions $\{Q_k\}_{k \geq 0}$ of \mathcal{X} such that $\text{card}(Q_0) = 1$ and $\text{card}(Q_k) = 2^{2^k}$ for $k \geq 1$. By a sequence of increasing partitions, we mean that every set in Q_k is contained in some set of Q_{k+1} . We will use $Q_k(X)$ to denote the unique set in the partition Q_k that contains $X \in \mathcal{X}$. The diameter of $Q_k(X)$ is denoted by $\Delta(Q_k(X))$. Then we have the following definition for γ_p functional associated with a metric space:

Definition 7 Suppose (\mathcal{X}, d) is a metric space and $p > 0$. We define

$$\gamma_p(\mathcal{X}, d) = \inf \sup_{X \in \mathcal{X}} \sum_{k \geq 0} 2^{k/p} \Delta(Q_k(X)), \quad (100)$$

where the infimum is taken over all admissible sequences.

The importance of the γ_p functional lies in its relationship with the behavior of a Gaussian process indexed by a metric space when the metric coincides with the one induced by the Gaussian process. More precisely, suppose $\{\xi_X\}_{X \in \mathcal{X}}$ is a Gaussian process indexed by the metric space (\mathcal{X}, d) with

$$d(X, Z) = (\mathbb{E}(\xi_X - \xi_Z)^2)^{1/2}, \quad (101)$$

then we have

$$c\gamma_2(\mathcal{X}, d) \leq \mathbb{E} \sup_{X \in \mathcal{X}} \xi_X \leq C\gamma_2(\mathcal{X}, d) \quad (102)$$

for some numerical constants c and C . The upper bound was first established by Fernique [33] and the lower bound is obtained by Talagrand using majorizing measures [34]. The rather difficult concept of majorizing measures has been considerably simplified through the notion of "generic

chaining”, an idea that dates back to Kolmogorov and is greatly advanced in recently years by Talagrand [24]. The upper bound (generic chaining bound)

$$\mathbb{E} \sup_{X \in \mathcal{X}} \xi_X \leq C \gamma_2(\mathcal{X}, d) \quad (103)$$

is actually applicable as long as the increments of $\{\xi_X\}_{X \in \mathcal{X}}$ have subgaussian tails:

$$\mathbb{P}\{|\xi_X - \xi_Z| > t\} \leq c \exp\left(-\frac{t^2}{2d(X, Z)^2}\right), \quad \forall t > 0 \quad (104)$$

for $d(X, Z)$ defined in (101) and

$$\mathbb{E} \xi_X = 0, \quad \forall X \in \mathcal{X}. \quad (105)$$

Under the conditions (104) and (105), an immediate consequence of the generic chaining bound is the well-known Dudley’s inequality [22], [24]

$$\mathbb{E} \sup_{X \in \mathcal{X}} \xi_X \leq C \sum_{k \geq 0} 2^{k/2} e_k(\mathcal{X}), \quad (106)$$

or equivalently in the more familiar integral form

$$\mathbb{E} \sup_{X \in \mathcal{X}} \xi_X \leq C \int_0^\infty \sqrt{\log N(\mathcal{X}, d, \epsilon)}, \quad (107)$$

where $e_k(\mathcal{X})$ and $N(\mathcal{X}, d, \epsilon)$ are the entropy number and covering number [31], [32], respectively. In general the generic chaining bound (103) is tighter than the Dudley’s entropy bounds (106) and (107).

We can not apply the generic chaining bound (103) directly to the random process (99) because this random process does not have increments with subgaussian tails, even if the entries of A_i in (99) are independent subgaussian. (The zero mean condition is a minor issue.) As a matter of fact, the increments of $\{\xi_X\}_{X \in \mathcal{X}}$ defined in (99) is a mixture of subgaussian and subexponential:

$$\mathbb{P}\{|\xi_X - \xi_Z| > t\} \leq 2 \exp\left(-\min\left(\frac{t^2}{d(X, Z)}, \frac{t}{d(X, Z)}\right)\right) \quad \forall X, Z \in \mathcal{X}, \quad t > 0. \quad (108)$$

In this case, the generic chaining bound becomes [24]

$$\mathbb{E} \sup_{X \in \mathcal{X}} \leq C(\gamma_1(\mathcal{X}, d) + \gamma_2(\mathcal{X}, d)). \quad (109)$$

The Dudley’s entropy bounds (106) and (107) also need to be modified to include an additional term corresponding to the subexponential tail. We remark that the $\gamma_1(\mathcal{X}, d)$ term and its counter

parts in Dudley's entropy inequality present major challenges of applying these bounds to our ℓ_* -CMSV problem. We need to circumvent this term in some way.

One major contribution of [23] is to show that the $\gamma_1(\mathcal{X}, d)$ term in (109) is not necessary for the empirical process $\{P_m(f^2)\}_{f \in \mathcal{F}_\tau}$. In particular, the expectation $\mathbb{E} \sup_{f \in \mathcal{F}_\tau} |P_m(f^2) - \mathbb{E} f^2|$ behaves as if the underlying process is subgaussian even if it is a mixture of subgaussian and subexponential. One key condition is, as commented by the authors of [23], all functions in \mathcal{F}_τ have the same second order moments. We now present the empirical process result:

Theorem 5 *Let $A \in \mathbb{R}^n$ be a random vector which induces a measure μ on \mathbb{R}^n , and \mathcal{F} be a subset of the unit sphere of $L_2(\mathbb{R}^n, \mu)$ with $\text{diam}(\mathcal{F}, \|\cdot\|_{\psi_2}) = \alpha$. Then there exist absolute constants c_1, c_2, c_3 such that for any $\epsilon > 0$ and $m \geq 1$ satisfying*

$$m \geq c_1 \frac{\alpha^2 \gamma_2^2(\mathcal{F}, \|\cdot\|_{\psi_2})}{\epsilon^2}, \quad (110)$$

with probability at least $1 - \exp(-c_2 \epsilon^2 m / \alpha^4)$,

$$\sup_{f \in \mathcal{F}} \left| \frac{1}{m} \sum_{k=1}^m f^2(A_k) - \mathbb{E} f^2(A) \right| \leq \epsilon. \quad (111)$$

Furthermore, if \mathcal{F} is symmetric, we have

$$\begin{aligned} & \mathbb{E} \sup_{f \in \mathcal{F}} \left| \frac{1}{m} \sum_{k=1}^m f^2(A_k) - \mathbb{E} f^2(A) \right| \\ & \leq c_3 \max \left\{ \alpha \frac{\gamma_2(\mathcal{F}, \|\cdot\|_{\psi_2})}{\sqrt{m}}, \frac{\gamma_2^2(\mathcal{F}, \|\cdot\|_{\psi_2})}{m} \right\}. \end{aligned} \quad (112)$$

As we remarked before, the key advantage of this theorem is that the results only involve $\gamma_2(\mathcal{F}, \|\cdot\|_{\psi_2})$. We now develop a proof of Theorem 1 based on Theorem 5.

Proof of Theorem 1: Consider the function set $\mathcal{F} = \mathcal{F}_\tau = \{f_X(\cdot) = \langle X, \cdot \rangle : \|X\|_F = 1, \|X\|_*^2 \leq \tau\}$. Assume $A \in \mathbb{R}^{n_1 \times n_2}$ and $\text{vec}(A)$ is isotropic and subgaussian with constant L . As a consequence of the isotropy of $\text{vec}(A)$ and $\|\text{vec}(X)\|_2 = \|X\|_F = 1$, we get \mathcal{F}_τ is a subset of the unit sphere of $L_2(\mathbb{R}^{n_1 n_2}, \mu)$. The symmetry of \mathcal{F}_τ yields

$$\begin{aligned} \alpha &= \text{diam}(\mathcal{F}_\tau, \|\cdot\|_{\psi_2}) \\ &= 2 \sup_{X \in \mathcal{H}_\tau} \|\langle X, A \rangle\|_{\psi_2} \\ &\leq 2 \sup_{X \in \mathcal{H}_\tau} L \|X\|_F \\ &= 2L. \end{aligned} \quad (113)$$

Now the key is to compute $\gamma_2(\mathcal{F}_\tau, \|\cdot\|_{\psi_2})$. Due to (102), the problem reduces to computing $\mathbb{E} \sup_{X \in \mathcal{H}_\tau} \zeta_X$ (actually an upper bound suffices), where $\{\zeta_X\}_{X \in \mathcal{H}_\tau}$ is the canonical Gaussian process on \mathcal{X} :

$$\zeta_X = \langle G, X \rangle, \text{vec}(G) \sim \mathcal{N}(0, I_{n_1 n_2}), X \in \mathcal{H}_\tau. \quad (114)$$

Clearly, we have

$$\begin{aligned} \gamma_2(\mathcal{F}_\tau, \|\cdot\|_{\psi_2}) &\leq c \mathbb{E} \sup_{X \in \mathcal{H}_\tau} \langle G, X \rangle \\ &\leq c \|X\|_* \mathbb{E} \|G\|_2 \\ &\leq c \sqrt{\tau} \sqrt{n_2}. \end{aligned} \quad (115)$$

As a consequence, the conclusions of Theorem 1 hold. ■

VI. CONCLUSIONS

In this paper, the ℓ_* -constrained minimal singular value of a measurement operator, which measures the invertibility of the measurement operator restricted to matrices with small ℓ_* -rank, is proposed to quantify the stability of low-rank matrix reconstruction. The reconstruction errors of the matrix Basis Pursuit, the matrix Dantzig selector, and the matrix LASSO estimator are concisely bounded using the ℓ_* -CMSV. Using a generic chaining bound for empirical processes, we demonstrate that the ℓ_* -CMSV is bounded away from zero with high probability for the subgaussian measurement ensembles, as long as the number of measurements is relatively large.

In the future work, we will study the feasibility of using the ℓ_* -CMSV to bound the reconstruction error of iterative algorithms. More importantly, we will also design algorithms to efficiently compute the ℓ_* -CMSV and use the ℓ_* -CMSV as a basis for designing optimal sensing operators.

REFERENCES

- [1] E. J. Candés and T. Tao, “Near-optimal signal recovery from random projections: Universal encoding strategies?,” *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, dec. 2006.
- [2] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [3] B. Recht, M. Fazel, and P. Parrilo, “Guaranteed minimum rank solutions of matrix equations via nuclear norm minimization,” Submitted to *SIAM Review*, 2007.
- [4] M. Fazel, *Matrix Rank Minimization with Applications*, Ph.D. thesis, Stanford University, 2002.
- [5] E. J. Candés and Y. Plan, “Tight oracle bounds for low-rank matrix recovery from a minimal number of random measurements,” Submitted for publication, 2009.

- [6] J. B. Tenenbaum, V. de Silva, and J. C. Langford, “A global geometric framework for nonlinear dimensionality reduction,” *Science*, vol. 290, no. 5500, Dec. 2000.
- [7] S. T. Roweis and L. K. Saul, “Nonlinear dimensionality reduction by locally linear embedding,” *Science*, pp. 2323–2326, Dec. 2000.
- [8] L. El Ghaoui and P. Gahinet, “Rank minimization under LMI constraints: A framework for output feedback problems,” in *Proceedings of the European Control Conference*. 1993.
- [9] M. Fazel, H. Hindi, and S. Boyd, “A rank minimization heuristic with application to minimum order system approximation,” in *Proceedings of the American Control Conference*. 2001.
- [10] E. J. Candès and B. Recht, “Exact matrix completion via convex optimization,” *Found. of Comput. Math.*, vol. 9, no. 7, pp. 717–772, 2008.
- [11] E. J. Candès and Y. Plan, “Matrix completion with noise,” Submitted to *Proceedings of the IEEE*.
- [12] D. Gross, Y. K. Liu, S. T. Flammia, S. Becker, and J. Eisert, “Quantum state tomography via compressed sensing,” 2009.
- [13] R. Basri and D.W. Jacobs, “Lambertian reflectance and linear subspaces,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 2, pp. 218 – 233, feb 2003.
- [14] E. J. Candès, X. Li, Y. Ma, and J. Wright, “Robust principal component analysis?,” Technical report, Stanford University, 2009.
- [15] N. Linial, E. London, and Y. Rabinovich, “The geometry of graphs and some of its algorithmic applications,” *Combinatorica*, vol. 15, pp. 215–245, 1995.
- [16] G. Tang and A. Nehorai, “The ℓ_1 -constrained minimal singular value: a computable quantification of the stability of sparse signal reconstruction,” Submitted to *IEEE Trans. Inf. Theory*.
- [17] A. Sofer S. G. Nash, *Linear and nonlinear programming*, McGraw-Hill, New York, NY, 1996.
- [18] W. W. Bradbury and R. Fletcher, “New iterative methods for solutions of the eigenproblem,” *Numer. Math.*, vol. 9, pp. 259–267, 1966.
- [19] H. Yang, “Conjugate gradient methods for the Rayleigh quotient minimization of generalized eigenvalue problems,” *Computing*, vol. 51, pp. 79–94, 1993.
- [20] S. Mendelson and N. Tomczak-Jaegermann, “A subgaussian embedding theorem,” *Israel Journal of Mathematics*, pp. 349–364, Mar. 2008.
- [21] F. Barthe, O. Guédon, S. Mendelson, and A. Naor, “A probabilistic approach to the geometry of the ℓ_p^n -ball,” *Ann. Probab.*, pp. 480–513, 2005.
- [22] M. Ledoux and M. Talagrand, *Probability in Banach spaces: Isoperimetry and processes*, vol. 23. of *Ergebnisse der Mathematik und ihrer Grenzgebiete*, 1991.
- [23] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, “Reconstruction and subgaussian operators in asymptotic geometric analysis,” *Geometric And Functional Analysis*, pp. 1248–1282, Nov. 2007.
- [24] M. Talagrand, *The generic chaining: upper and lower bounds of stochastic processes*, Springer, 2005.
- [25] S. Ma, D. Goldfarb, and L. Chen, “Fixed point and bregman iterative methods for matrix rank minimization,” *Mathematical Programming*, 2009.
- [26] E. J. Candès, “The restricted isometry property and its implications for compressed sensing,” *Compte Rendus de l’Academie des Sciences, Paris, Serie I*, vol. 346, 2008.
- [27] P. Bickel, Y. Ritov, and A. Tsybakov, “Simultaneous analysis of Lasso and Dantzig selector,” *Annals of Statistics*, vol. 37, no. 4, pp. 1705–1732, 2009.

- [28] K. Davidson and S. Szarek, “Local operator theory, random matrices and banach spaces,” in *Handbook of the Geometry of Banach Spaces*, W. B. Johnson and J. Lindenstrauss, Eds., vol. 1, pp. 317–366. North Holland, New York, NY, 2001.
- [29] R. Vershynin, *Lecture notes on non-asymptotic random matrix theory*, 2007.
- [30] Michel Talagrand, “Spin glasses: a new direction for probability theory?,” *Atti Accad. Naz. Lincei Cl. Sc. Fis. Math. Nat. Rend. Lincei Mat. Appl.*, special issue, pp. 127–146, 2000.
- [31] B. Carl and I. Stephani, *Entropy, compactness, and the approximation of operators*, Cambridge University Press, 1990.
- [32] David Eric Edmunds and Hans Triebel, *Function spaces, entropy numbers, differential operators*, Cambridge University Press, Jan. 1996.
- [33] X. Fernique, *Régularité des trajectoires des fonctiones aléatoires gaussiennes*, Ecole d’Eté de Probabilités de St-Flour 1974, Lecture Notes in Mathematics 480, Springer-Verlag, 1975.
- [34] M. Talagrand, “Regularity of Gaussian processes,” *Acta Math.*, vol. 159, pp. 99–149, 1987.